



Abordagens de Text Mining (TM)

Bem-vindo ao nosso estudo sobre abordagens de Text Mining - TM. Este estudo ajudará você a compreender melhor como estudar a disciplina de Desenvolvimento de Algoritmos de Inteligência Artificial. Neste módulo será apresentado um exemplo de um projeto de Text Mining e suas etapas na prática, a ideia é simular o projeto de desenvolvimento de um software para descobrir novos conhecimentos em textos de fichas de atendimento telefônico, utilizando para isso técnicas de mineração em texto. Conhecer os conceitos que serão apresentados aqui te ajudará a chegar ao sucesso do seu estudo. Vamos começar a estudar?

Metodologia

Segundo o dicionário Oxford, metodologia é "ramo da lógica que se ocupa dos métodos das diferentes ciências". É possível dizer que, são os passos para desenvolver uma pesquisa ou projeto, como foi desenvolvido para chegar em um resultado.

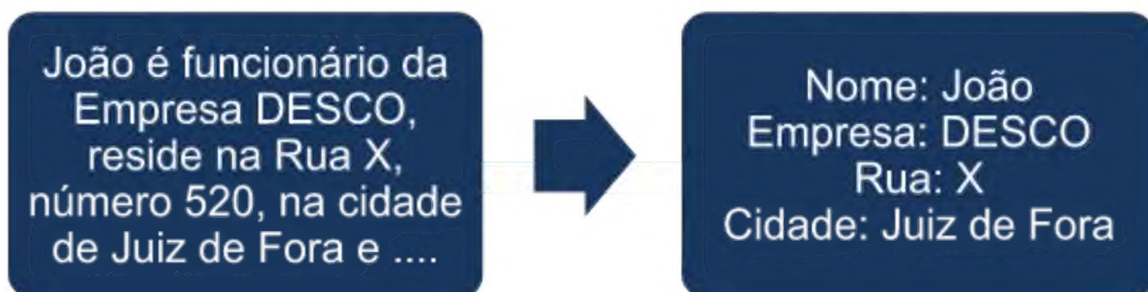
Então para se desenvolver um projeto, é necessário ter uma metodologia. Por exemplo:

- Definição de quais informações serão extraídas e de onde: e-mails, textos livres, arquivos eletrônicos, entre outros. No exemplo desse módulo, a extração de informações será feita em fichas de atendimento telefônico.
- Quais as etapas ou passos para chegar no objetivo. Exemplo: Compreensão do negócio, nesse projeto são as fichas de atendimento que são os registros relevantes para a análise. Compreensão dos dados, etapa de analisar as palavras e extrair as que são relevantes. Preparação dos dados, sua principal função é fornecer um primeiro nível de informação textual estruturada, como o reconhecimento de início e fim de sentenças e classificação de palavras quanto a sua função sintática (limpeza dos dados). Modelagem e avaliação

dos algoritmos de Text Mining escolhidos para a mineração. Avaliação do projeto como um todo e revisão dos passos seguidos. E por fim, se a aplicação atingiu o objetivo (registros classificados corretamente).

Técnica

Nessa disciplina já foram apresentadas algumas etapas e técnicas de Text Mining, tais como: Recuperação da Informação (Coleta e Pré-processamento), seu objetivo é localizar os documentos que contém informações definidas pelo usuário em uma consulta. Para agilizar, utiliza-se a indexação, extraindo assim os termos mais significativos e excluindo os que não têm importância. Extração das Informações (Mineração), sua função é encontrar valores implícitos nos textos (Figura 1).



Desenvolvimento

Esta é a etapa de definir o que será implementado no novo software para resolver o problema estudado e informado pelos atendentes do suporte e pela equipe técnica da empresa, definido no início deste módulo. Atualmente, o cliente utiliza um software que serve para cadastramento da ficha de atendimento telefônico em texto corrido (documentos que serão selecionados para mineração de texto). Os problemas relatados pelos clientes e os erros encontrados pela própria equipe técnica são descritos na documentação do software antes do mesmo ser desenvolvidos, tais como:

1) Requisitos não funcionais do sistema:



- Desempenho

- Banco de dados

- Visualização

2) Requisitos funcionais (que o sistema deverá fazer):

- Listar palavras.

- Listar palavras excluídas (stopwords)

- Criar e remover categorias;

- Listar palavras chaves (keywords).

- Listar frequência.

- Criar gráficos.

- Buscar registros por palavras.

- Buscar frases.

Especificação e Implementação

Seguindo a metodologia escolhida, o estudo foi feito com o foco nas fichas de atendimento telefônico. Para isto foram analisadas todas as tabelas envolvidas nas fichas.

Os dados analisados estavam claros e organizados, sem erros de grafia, portanto a etapa de preparação dos dados não foi realizada. A modelagem foi realizada com a ajuda de ferramentas de Engenharia

de Software, como por exemplo: Jude Community para a especificação do Diagrama de Caso de Uso, e Enterprise Architect para especificar o Diagrama de Classes.

Na etapa de avaliação da modelagem foi possível verificar que o modelo estava adequado e eficiente apontando as fichas de atendimento que estavam com a situação incorreta em relação ao texto contido na descrição. Por fim, a implementação do software em uma linguagem de programação desenvolvida de acordo com toda documentação criada no projeto.

Após a implementação do software, o mesmo deverá ser testado e uma verificação dos resultados é de suma importância para verificar se o problema proposto foi resolvido, e posteriormente implantar o software na empresa.

Neste módulo foi apresentado um resumo de um exemplo de projeto de um software de Text Mining com o objetivo de minerar informações relevantes e corrigir, caso necessário, textos de fichas de atendimentos de uma empresa. Alguns detalhes foram abstraídos pois o foco está na mineração de textos e não em engenharia de software.

Atividade Extra

Para se aprofundar no assunto desta aula assista o filme “Ron Bugado” (2021)

Referência Bibliográfica

ARANHA, C. **Uma abordagem de pré-processamento automático para mineração de textos em português: sob o enfoque da inteligência computacional**. Departamento de Engenharia Elétrica, PUC-Rio. Rio de Janeiro. 2007.

FILHO, J. A. C. **Mineração de Textos: Análise de Sentimentos Utilizando Tweets Referentes à Copa do Mundo de 2014**. Universidade Federal do Ceará. Quixadá. 2014.

HESSLER, J. DE C. **Um método para apoio à leitura baseado no uso de uma ferramenta de mineração de texto**. Universidade Federal do Rio Grande do Sul. Porto Alegre. 2010.

Morais, E. A. M. e Ambrósio, A. P. L. **Mineração de Textos**. Universidade Federal de Goiás. Goiás. 2007.

PUC-Rio. **Mineração de Textos: Fundamentos e Aplicações**. Rio de Janeiro. Disponível em <https://www.maxwell.vrac.puc-rio.br/11675/11675_3.PDF> Acesso em: 04 de Outubro de 2022.

VALDATI, Aline de Brittos. **Inteligência Artificial – IA**. São Paulo: Contentus; 2020

Ir para exercício